친지인 최종 Proposal

(Chin-ji-in, 櫬地人)



오픈소스SW프로젝트 01분반 – TEAM 9

이창민 서정현 한태균 임우섭 정석우

# 목차

1. **서론**
   1. 프로젝트 배경
      1. 친지인이란?
      2. 천지인 키보드 특징
      3. 천지인 오타 특징
   2. 기존 기술 조사
      1. 스마트폰에 내장된 자동완성 기능
      2. Peter Norvig
      3. Symspell
      4. Customed-Symspell
      5. Hanspell
2. **본론**
   1. 개발 필요성
      1. 천지인 오타 통계
      2. 천지인 오타와 두벌식 오타의 차이
   2. 개발 방법
      1. 개발 언어 및 도구
      2. 방법론?
      3. 알고리즘
3. **결론**
   1. 프로젝트 의의

# 프로젝트 배경

## 친지인(櫬地人)

천지인 키보드에서 오타를 줄이기 위해 고안된 프로젝트이다. 단어 ‘천지인’에서 아래아(ㆍ)가 탈락된 오타로 이름이 지어졌으며, 한자는 무궁화 나무 친 ‘櫬’을 사용하여 한글 창제 원리를 기반으로 한 천지인 키보드가 우리나라 사람들에게 보다 쉽게 사용될 수 있길 바라는 마음을 담았다.

## 천지인 키보드 특징

가장 흔히 사용되는 두벌식(qwerty) 키보드와 비교해보면 천지인 키보드는 기본적으로 필요한 자판이 문자 자판 10개와 공백 자판으로 총 11개인 반면, 두벌식 키보드는 기본적으로 문자 자판 26개와 공백 자판, Shift 자판이 필요하므로 총 28개의 자판을 가진다. 따라서 천지인 키보드는 자판이 넓지만 타수가 많아지고, 두벌식 키보드는 자판은 좁지만 타수는 적어진다. 이에 천지인은 두벌식 보다 한 손으로 입력하기 용이하다는 장점을 가진다.

많은 타수를 입력해야 하는 천지인 키보드 사용자는 두벌식 키보드 사용자보다 빠른 입력이 요구되기 때문에 넓은 자판에도 불구하고 자판을 잘못 누르는 경우가 종종 발생한다. 특히 자판 간의 순서 바뀜, 자판 하나 안눌림, 자판 하나 더눌림과 같은 오타는 자판의 크기와 크게 관련없이 발생하기 때문에 **천지인에서의 오타가 두벌식에서보다 빈번히 발생**하는 것으로 조사되었다.

## 천지인 오타 특징

천지인에서의 오타 결과는 두벌식과 차이가 크다. 예를 들어, ‘고양이’라는 단어에서 양과 이 사이에 스페이스바를 누르지 않으면 ‘고야미’라는 결과가 나오고, 양에서 아래아(ㆍ) 하나를 ‘ㅣ’보다 먼저 누르면 ‘고어 ㆍㅇ이’ 라는 결과가 나오게 되는데 이는 모두 두벌식에서는 나타나기 힘든 오타이다. 따라서 천지인 전용의 오타 교정 알고리즘이 필요하다.

# 기존 기술 조사

스마트폰에서 쉽게 볼 수 있는 자동 완성 기능과, 영어 오타 교정 알고리즘으로 유명한 **Peter-Norvig Algorithm**, **symspell** 알고리즘을 조사하고, symspell에 두벌식 키보드 자판의 물리적 거리 가중치를 적용한 **customed-symspell**과, symspell을 한글에 적용한 **py-hanspell** 프로젝트를 조사하였다.

## 자동 완성 기능

기존에 알고 있던 단어와 사용자가 자주 입력했던 단어를 기반으로, 현재 입력중인 단어를 완성또는 대체할 단어들을 추천해준다.

사용자 입력 단어를 학습한다는 것은 큰 장점이지만, 오타교정 목적으로 사용하기에는 뚜렷한 한계가 존재한다. ‘고양미’나 ‘고야이’ 같은 경우는 고양이로 고쳐주지만 ‘노양이’는 못 고쳐주는 것으로 보아 이 기능은 오타교정보다는 자동완성을 목적을 두고 있기에 단어 초반의 오타를 잡아주지 못하는 것으로 보인다. 또한 위에서 언급했던 특수한 천지인 오타(‘고야미’, ‘고어 ㆍㅇ이’ 등)는 전혀 잡지 못하는 것으로 보아 **‘친지인’ 프로젝트와는 차이가 있다.**

Peter Norvig 알고리즘

Peter Norvig 알고리즘은 미국의 컴퓨터 과학자 Peter Norvig에 의해 고안된 가장 대표적인 오타 교정 알고리즘이다.